Crystallography Open Database for teaching

Saulius Gražulis Andrius Merkys Antanas Vaitkus

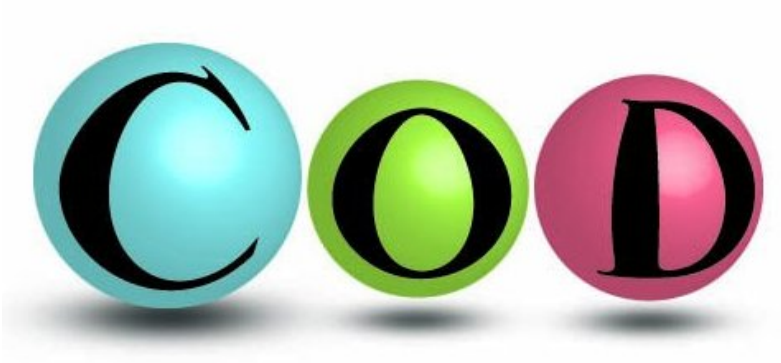
Hyderabad, 24th IUCr Congress, 2017

Vilnius University Institute of Biotechnology

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Crystallography Open Database

Largest open-access resource on chemical crystallography



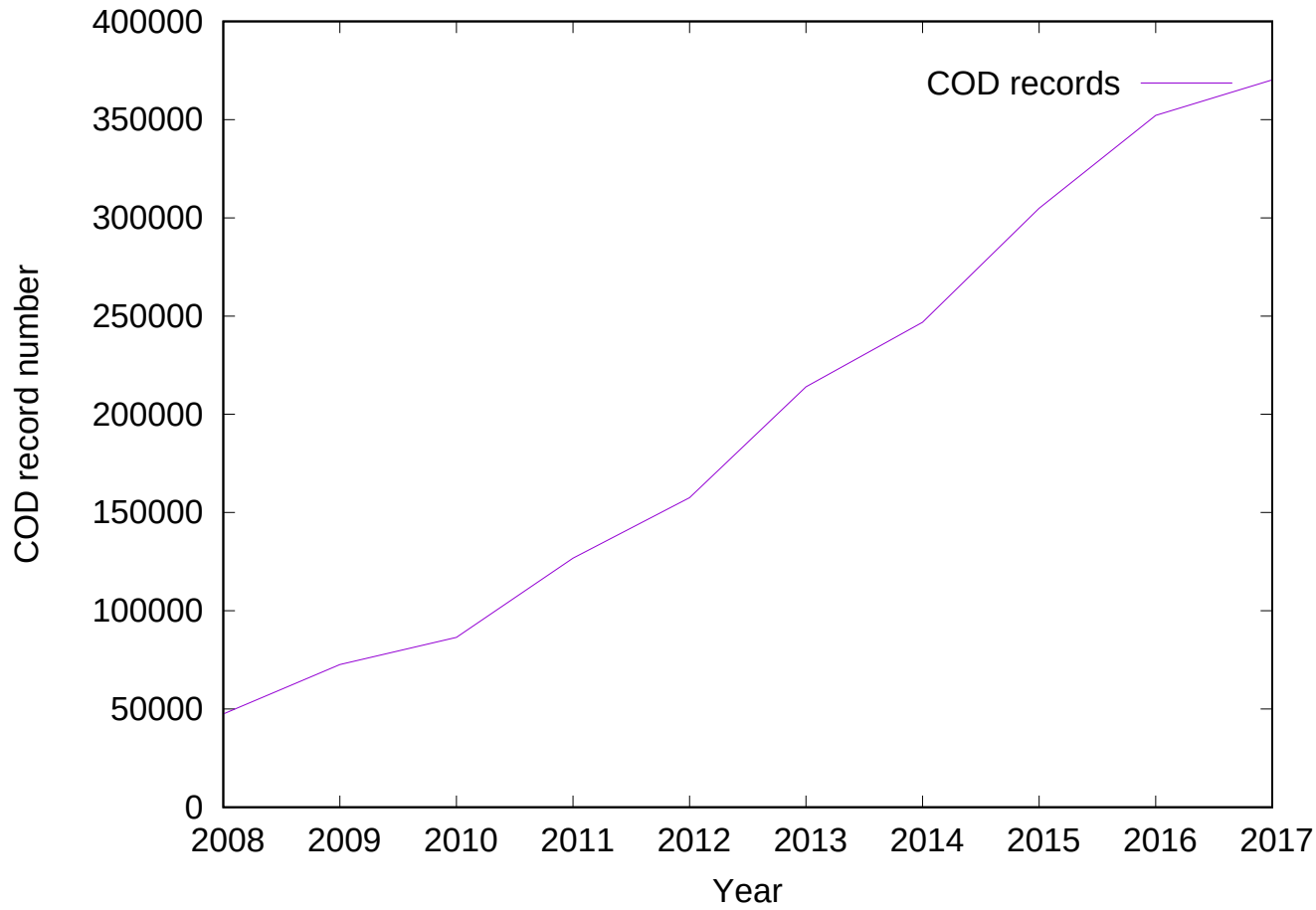
http://www.crystallography.net/

- Total >**380 000** records;
- Grows about 30-50 thous. records/year;

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

COD persistence

COD is on-line since 2003, growing all the time.



This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

COD accessibility

COD is a **fully open-access database**. All records are available under public domain designation.

Every crystallographer can contribute data to COD (and many do ;).

Provided access methods are:

- Web search
- URLs constructed from stable identifiers
- RESTful interfaces
- Full data download

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Use of COD for teaching crystallography

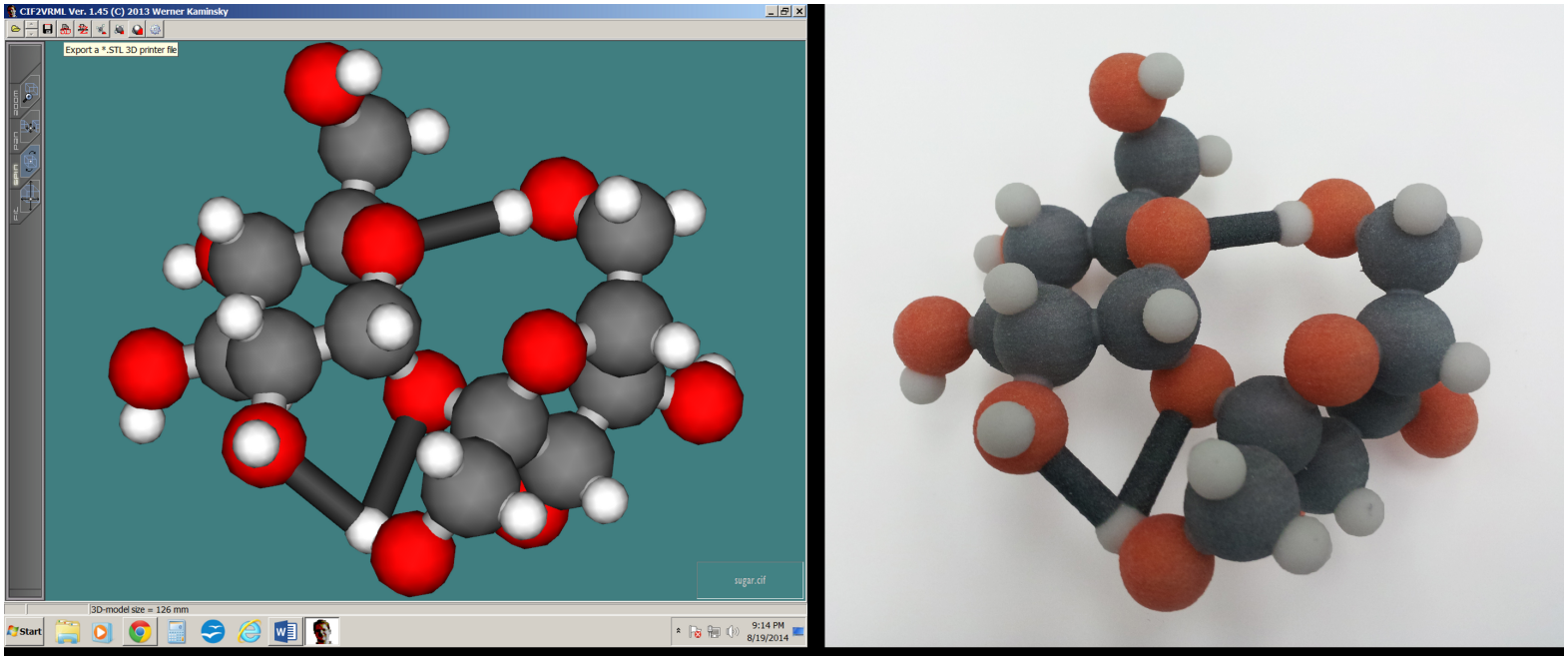
We can and may use COD for teaching in following ways:

- as a source of crystallographic information – to illustrate concepts of crystallography;
- as a programmatically accessible resource – to teach data processing and management skills on real-life sized repositories;
- as an goal in itself – students can participate in COD co-development;

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

3D printing of models

Use COD to produce 3D Virtual reality descriptions and to eventually make 3D printed models [Kaminsky et al., 2014, Gražulis et al., 2015]:



This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

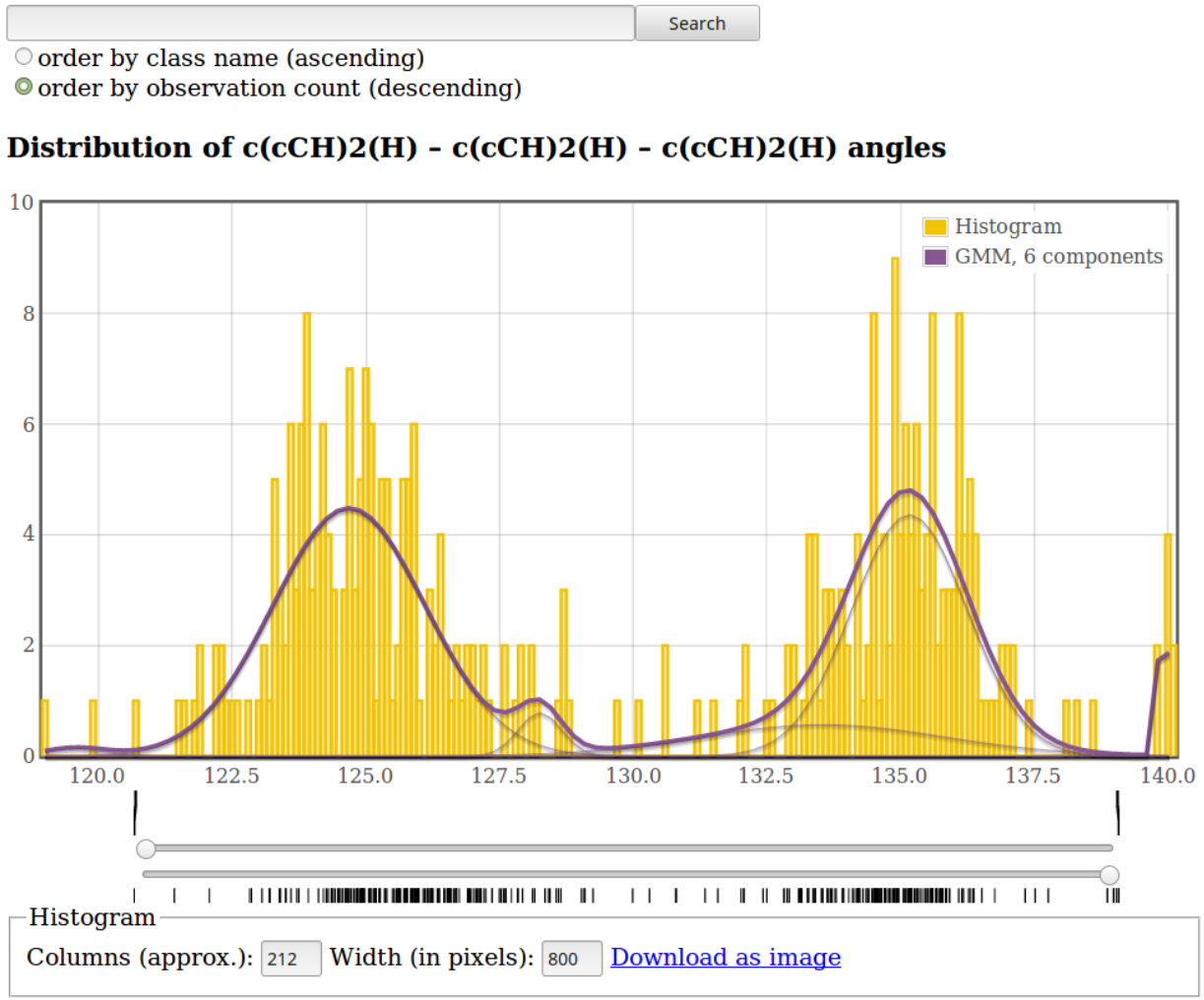
Personal story: Andrius Merkys

- Defended his BSc in Bioinformatics (Vilnius University), designed and implemented data deposition interface for the Crystallography Open Database;
- Defended master thesis with honours (Vilnius University), providing the Crystallography Open Database as a source of data for MM model refinement;
- Part of results published together with the group of Garib Murshudov [Long et al., 2017];

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Models in the COD geometry browser

http://www.crystallography.net/geometry/cgi-bin/histogram.pl?class_id=3313-3313-3313



This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

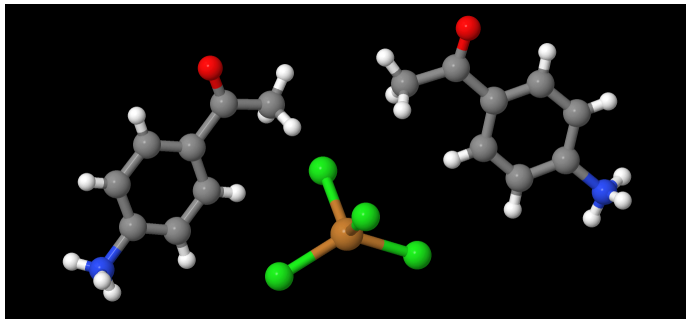
Personal story: Antanas Vaitkus

- Defended his BSc in Bioinformatics (Vilnius University), on the analysis of interatomic distance distributions in the COD;
- Defended his master thesis (with highest evaluation) on the analysis of metal coordination sphere geometries in COD;
- Works on the PhD thesis; topic: extraction of chemical knowledge from open Web resources and linking them with the COD;

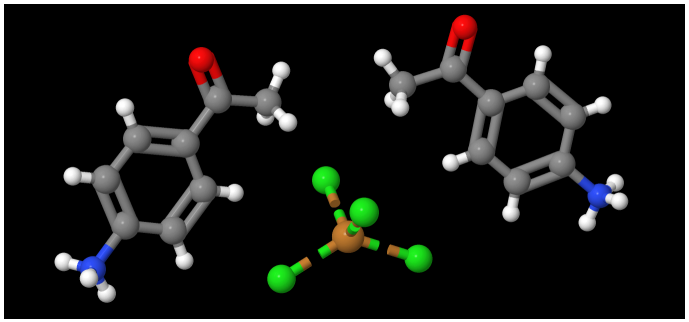
This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Comprehension of chemical information

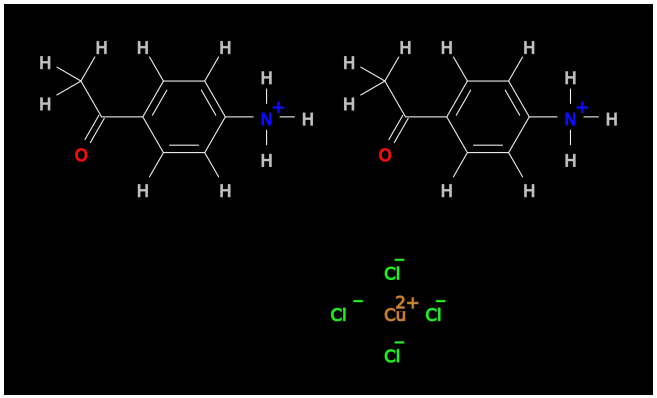
cif_molecule



cif2molecule



Open Babel



[Cu+2]([Cl-])([Cl-])([Cl-])[Cl-].O=C(c1ccc(cc1)[NH3+])C.O=C(c1ccc(cc1)[NH3+])C

Work done with Thomas Sander using his F/LOSS libraries at Actelion

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

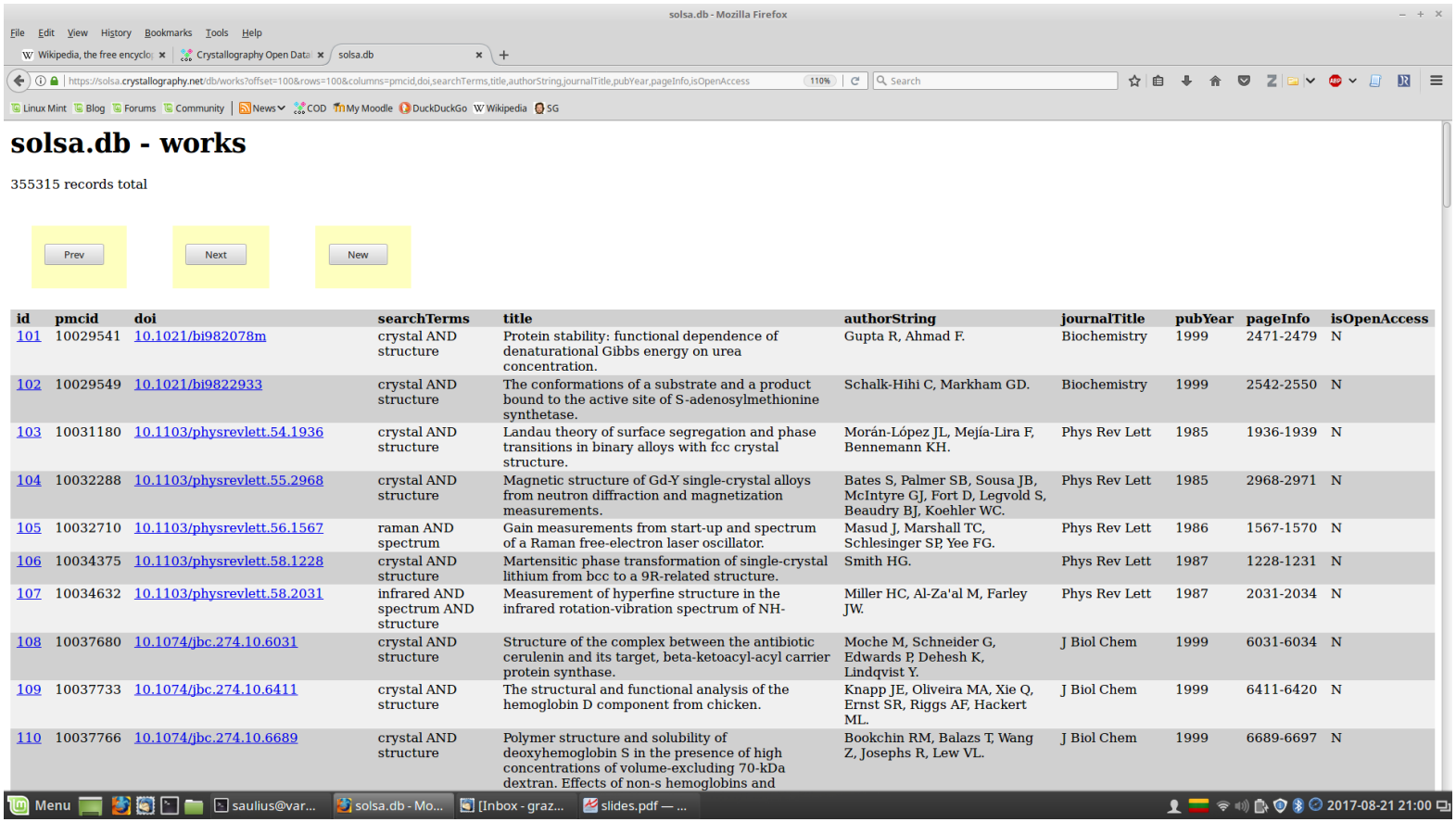
Personal story: Mélanie Lailler

- Engineer in training at ISTP, in Saint-Etienne (France);
- Visited Vilnius University for 2 months as a part of her internship program;
- In the 2 months, using the open EuropePMC resource, covered software development topics:
  - version control (Subversion);
  - software testing using Make, unite tests, test coverage;
  - software release cycle management;
  - wrote and released a finished program that is planned to be used further for data management int COD and ROD;

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Raman spectra and mineral structures in literature

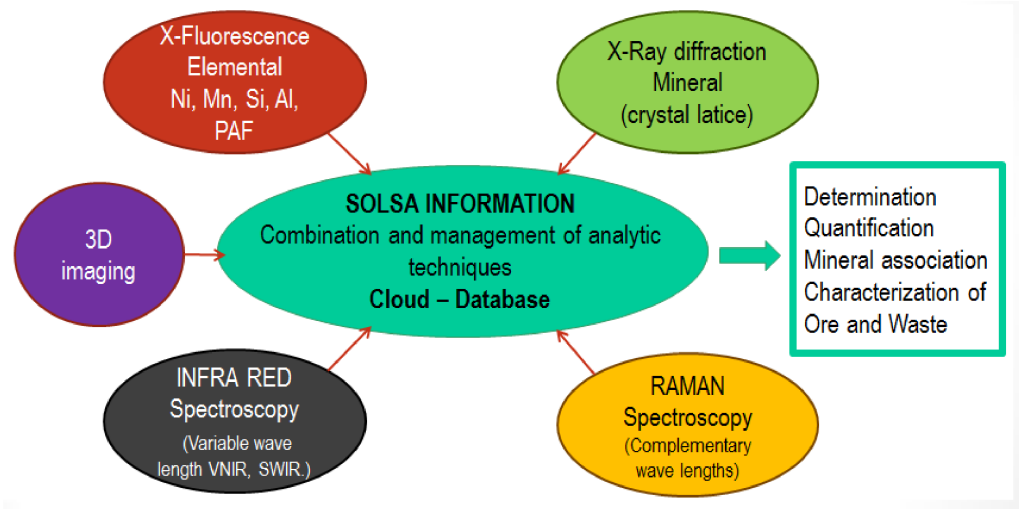
Automated searches in the open-access EuropePMC resource:



id	pmcid	doi	searchTerms	title	authorString	journalTitle	pubYear	pageInfo	isOpenAccess
101	10029541	10.1021/bi982078m	crystal AND structure	Protein stability: functional dependence of denaturational Gibbs energy on urea concentration.	Gupta R, Ahmad F.	Biochemistry	1999	2471-2479	N
102	10029549	10.1021/bi9822933	crystal AND structure	The conformations of a substrate and a product bound to the active site of S-adenosylmethionine synthetase.	Schalk-Hihi C, Markham GD.	Biochemistry	1999	2542-2550	N
103	10031180	10.1103/physrevlett.54.1936	crystal AND structure	Landau theory of surface segregation and phase transitions in binary alloys with fcc crystal structure.	Morán-López JL, Mejía-Lira F, Bennemann KH.	Phys Rev Lett	1985	1936-1939	N
104	10032288	10.1103/physrevlett.55.2968	crystal AND structure	Magnetic structure of Gd-Y single-crystal alloys from neutron diffraction and magnetization measurements.	Bates S, Palmer SB, Sousa JB, McIntyre GJ, Fort D, Legvold S, Beaudry BJ, Koehler WC.	Phys Rev Lett	1985	2968-2971	N
105	10032710	10.1103/physrevlett.56.1567	raman AND spectrum	Gain measurements from start-up and spectrum of a Raman free-electron laser oscillator.	Masud J, Marshall TC, Schlesinger SP, Yee FG.	Phys Rev Lett	1986	1567-1570	N
106	10034375	10.1103/physrevlett.58.1228	crystal AND structure	Martensitic phase transformation of single-crystal lithium from bcc to a 9R-related structure.	Smith HG.	Phys Rev Lett	1987	1228-1231	N
107	10034632	10.1103/physrevlett.58.2031	infrared AND spectrum AND structure	Measurement of hyperfine structure in the infrared rotation-vibration spectrum of NH-	Miller HC, Al-Za'al M, Farley JW.	Phys Rev Lett	1987	2031-2034	N
108	10037680	10.1074/jbc.274.10.6031	crystal AND structure	Structure of the complex between the antibiotic cerulenin and its target, beta-ketoacyl-acyl carrier protein synthase.	Moche M, Schneider G, Edwards P, Dehesh K, Lindqvist Y.	J Biol Chem	1999	6031-6034	N
109	10037733	10.1074/jbc.274.10.6411	crystal AND structure	The structural and functional analysis of the hemoglobin D component from chicken.	Knapp JE, Oliveira MA, Xie Q, Ernst SR, Riggs AF, Hackert ML.	J Biol Chem	1999	6411-6420	N
110	10037766	10.1074/jbc.274.10.6689	crystal AND structure	Polymer structure and solubility of deoxyhemoglobin S in the presence of high concentrations of volume-excluding 70-kDa dextran. Effects of non-s hemoglobins and	Bookchin RM, Balazs T, Wang Z, Josephs R, Lew VL.	J Biol Chem	1999	6689-6697	N

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

SOLSA project and COD



COD will be used in SOLSA for:

- mineral identification;
- subsequent data dissemination.

SOLSA data flow diagram courtesy Monique Le Guen, ERAMET.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Contributions of students

- A.V.: design of the CIF dictionaries for Raman, IR, Hyperspectral image databases;
- A.M.: launching and curation of the ROD, HOD and spectral databases;
- M.L.: performed a literature search for potential publications with Raman spectra and crystal structures useful for mineral identification;

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Acknowledgements

VU Institute of Biotechnology

Virginijus Siksnys
(*head of the dept.*)

Andrius Merkys
Antanas Vaitkus

COD Advisory board

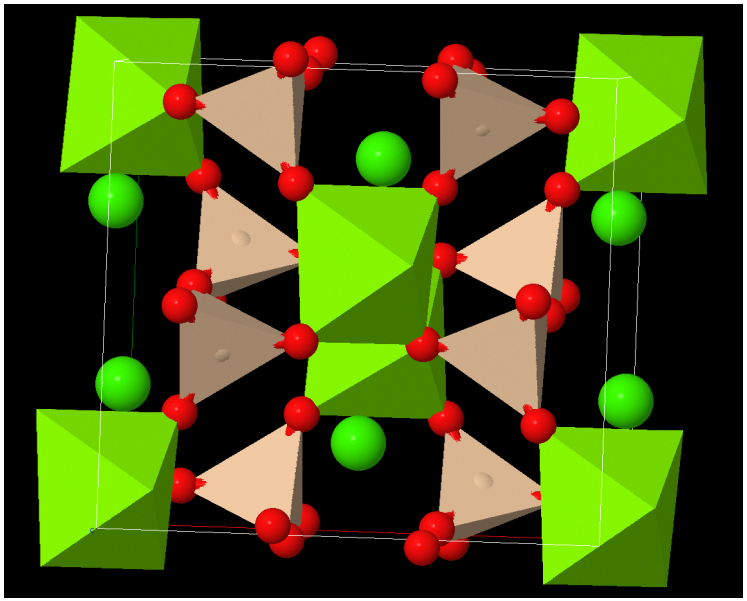
Daniel Chateigner
Robert T. Downs
Werner Kaminsky
Armel Le Bail
Luca Lutterotti
Peter Moeck
Peter Murray-Rust
Miguel Quirós

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

Thank you!

Rob Lavinsky, iRocks.com – CC-BY-SA-3.0,
Diopside-172005 CC BY-SA 3.0



http://www.crystallography.net/1000007.html

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.

References

Gražulis, S., Sarjeant, A. A., Moeck, P., Stone-Sundberg, J., Snyder, T. J., Kaminsky, W., Oliver, A. G., Stern, C. L., Dawe, L. N., Rychkov, D. A., Losev, E. A., Boldyreva, E. V., Tanski, J. M., Bernstein, J., Rabeh, W. M., and Kantardjieff, K. A. (2015).
Crystallographic education in the 21st century.
Journal of Applied Crystallography, 48(6):1964–1975.

Kaminsky, W., Snyder, T., Stone-Sundberg, J., and Moeck, P. (2014).
One-click preparation of 3d print files (*.stl, *.wrl) from *.cif (crystallographic information framework) data using cif2vrml.
Powder Diffraction, 29:S42–S47.

Long, F., Nicholls, R. A., Emsley, P., Gražulis, S., Merkys, A., Vaitkus, A., and Murshudov, G. N. (2017).
ACEDRG: A stereo-chemical description generator for ligands.
Acta Crystallographica Section D, 73(2):112–122.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 689868.